



**HAL**  
open science

# La Banque de données du CDSP et la valorisation des Enquêtes électorales

Jieun Jeong, Lucie Marie

► **To cite this version:**

Jieun Jeong, Lucie Marie. La Banque de données du CDSP et la valorisation des Enquêtes électorales. Ouverture des données : l'Enquête électorale 2017 du CEVIPOF publiée par le CDSP, Centre de données socio-politiques de Sciences Po (Sciences Po, CNRS), Dec 2024, Paris, France. 10.5281/zenodo.14258132 . hal-04834217

**HAL Id: hal-04834217**

**<https://sciencespo.hal.science/hal-04834217v1>**

Submitted on 12 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

---

# La Banque de données du CDSP & la valorisation des Enquêtes électorales

Jieun JEONG, Lucie MARIE - Sciences Po, Centre de données socio-politiques (CDSP), CNRS

*Demi-journée dédiée à l'ouverture des données d'enquêtes électorales (CEVIPOF/CDSP)  
9 décembre 2024*

---

# Programme

## La Banque de données du CDSP

- Trouver
  - Commander
  - Déposer
- } des données

## La collection “Enquêtes électorales”

- Panorama des données
- Focus sur l’ENEF 2017
- Curation et enrichissement

## Agenda de l’équipe Documentation-Diffusion

- Prochaines publications de données
- Projets en cours (ANR, OSCAR)



CDSP'S  
DATABANK



**CTS certified collection with more than 400 references**  
ELIPSS longitudinal surveys, French Election Surveys, Electoral results since 1957

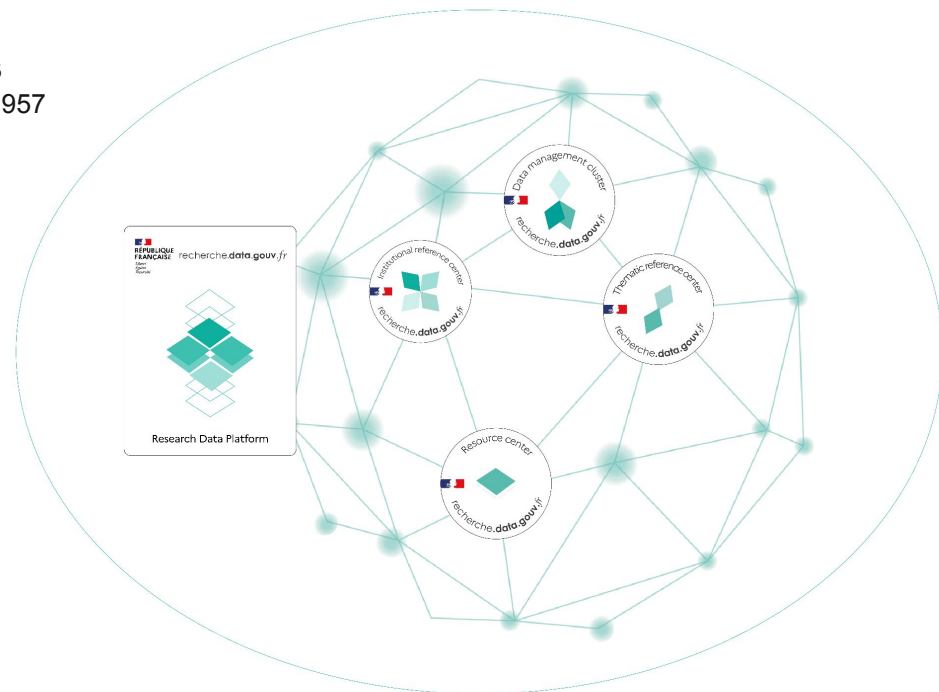
Open file formats, international metadata standards



Available on [data.sciencespo.fr/cdsp](https://data.sciencespo.fr/cdsp)



The  
**Dataverse**<sup>®</sup>  
Project



## Banque de données du CDSP

(Sciences Po, Centre de données socio-politiques (CDSP), CNRS)

[data.sciencespo](#) >

[Contact](#) [Partager](#)



Résultats électoraux



Bases de données  
pédagogiques



ELIPSS



Données qualitatives



Chercher dans ce dataverse...



[Recherche avancée](#)

**Dataverses (27)**

**Ensembles de données (437)**

**Fichiers (3 267)**

### Dataverse Category

Projet de recherche (14)

Laboratoire (3)

Groupe de recherche (1)

Chercheur (1)

Cours (1)

### Publication Year

2024 (43)

2023 (42)

2022 (49)

2021 (9)

1 à 10 de 464 résultats

Tri

**ELIPSS : Enquêtes doctorant-e-s** (Sciences Po Paris)



26 nov. 2024 ELIPSS

En 2024, le Centre de données socio-politiques a lancé un appel à projets spécifique destiné à soutenir les doctorant-e-s dans leur travail de thèse en sciences humaines et sociales. Ainsi, les jeunes chercheur-e-s sélectionné-e-s ont eu la possibilité d'administrer un module de...

**Santé Orale : Usages, Représentations et Inégalités de Recours (ELIPSS 2024)**



21 nov. 2024 - ELIPSS



Recherche Data Govv Génération datapaper

(Recherche Data Govv)

Statistiques

1 600 258 téléchargements

Contact Partager

Chercher dans cette collection...



Recherche avancée

Collections (0)

Jeux de données (50)

Fichiers (0)

Type de données

Text (50)

Audiovisuel (1)

Image (1)

Auteur Nom

Sciences Po, Centre de données socio-politiques (CDSP), CNRS (15)

Boy, Daniel (8)

Cautrès, Bruno (6)

Sciences Po, Centre de recherches politiques (CEVIPOF), CNRS (6)

Perrineau, Pascal (5)

Plus...

Mot-clé Terme

Social Sciences (50)

1 à 10 de 50 résultats

Tri ▾

Enquête post-électorale de l'élection présidentielle 2012

27 janv. 2024 - Institut d'études politiques de Paris



Sciences Po, Centre de recherches politiques (CEVIPOF), CNRS, 2014, "Enquête post-électorale de l'élection présidentielle 2012", <https://doi.org/10.21410/7E4/BNPAXT>

... Le dispositif d'enquêtes post-électorales de 2012 du Centre de recherches politiques de Sciences Po (CEVIPOF) a fait l'objet d'une expérimentation méthodologique. L'objectif était de se rapprocher de ce qui avait été réalisé lors des élections de 2007. En effet, il s'agissait d'évaluer les effets des modes d'administration du questionnaire sur la manière de répondre des enquêtés. ...

Auteur Nom: Sciences Po, Centre de recherches politiques (CEVIPOF), CNRS

Ce jeu de données est moisonné auprès de notre partenaire data.sciencespo.fr. En cliquant sur le lien, vous accédez directement au dépôt source des données.



Image de la science, 1989

27 janv. 2024 - Institut d'études politiques de Paris



Sciences Po, Centre de recherches politiques (CEVIPOF), CNRS, 2007, "Image de la science, 1989", <https://doi.org/10.21410/7E4/C3GIGY>

... En 1972, le Ministère de la Recherche a confié la réalisation d'une étude sur les attitudes du public à l'égard de la recherche scientifique française à une équipe de chercheurs du CEVIPOF. Cette recherche a donné lieu à une série de cinq enquêtes réalisées entre 1972 et 2001. Elle s'intéresse aux représentations de la science, à la perception du développement technique et économique et à l'opinion sur la politique de recherche. ...



Reset filters

Clear search

24 studies found in French from a total of 37720

About

User Guide

REST API

Topic



Keywords



Collection years



Country



Publisher



Results per page

30

Sort by

Relevance



1



### Enquête électorale française 2017 (ENEF 2017), Vague 9

Sciences Po, Centre de recherches politiques de Sciences Po (CEVIPOF), CNRS

Ce dispositif de suivi et de mesure de l'opinion permet l'analyse des évolutions de l'électorat français et de ses réactions aux événements politiques, sociaux et économiques à travers 20 vagues, depuis la veille des élections régionales de 2015 jusqu'au lendemain des élections législatives de 2017. Il permet des analyses détaillées au niveau national et régional pour saisir...

ATTITUDES POLITIQUES

DONNÉES ÉLECTORALES (ELSST)

Sciences sociales (ELSST)

OPINION PUBLIQUE (ELSST)

 Read more

 Study description available in: FR [Access study](#)

### Enquête électorale française 2017 (ENEF 2017), Vague 1Bis

Sciences Po, Centre de recherches politiques de Sciences Po (CEVIPOF), CNRS

Ce dispositif de suivi et de mesure de l'opinion permet l'analyse des évolutions de l'électorat français et de ses réactions aux événements politiques, sociaux et économiques à travers 20 vagues, depuis la veille des élections régionales de 2015 jusqu'au lendemain des élections législatives de 2017. Il permet des analyses détaillées au niveau national et régional pour saisir...

DONNÉES ÉLECTORALES (ELSST)

OPINION PUBLIQUE (ELSST)

ATTITUDES POLITIQUES

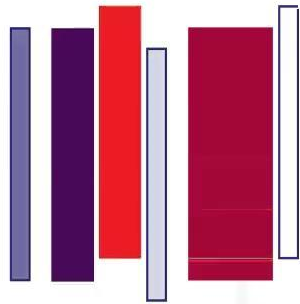
Sciences sociales (ELSST)

 Read more

 Study description available in: FR [Access study](#)

### Enquête électorale française 2017 (ENEF 2017), Vague 16

Send feedback



■

■



---

# Commander des données

## Procédure disponible sur le site du CDSP

Données > Commander des données

<https://www.sciencespo.fr/cdsp/fr/donnees/commander-des-donnees/>

Authentication et acceptation des conditions d'utilisation de l'entrepôt data.sciencespo.fr

## Tutoriels vidéos (bientôt disponibles)

- Se créer un compte
- Télécharger des données



Se créer un compte

# Déposer des données

Procédure disponible sur le site du CDSP

Données > Déposer des données

<https://www.sciencespo.fr/cdsp/fr/donnees/comman-der-des-donnees/>



CDSP

## Fiche de proposition d'enquête

### Fiche descriptive de l'enquête

Merci de remplir le plus précisément possible la fiche ci-dessous en fonction des matériaux déposés, en particulier les champs obligatoires (en gras).

Pour vous aider, une fiche complétée d'exemples est téléchargeable sur le site [du CDSP](#).

#### A- l'enquête

<b>A1</b>	<b>Titre de l'enquête</b> Lorsque l'enquête n'a pas de titre « officiel », préciser le titre de la publication principale à laquelle l'enquête a donné lieu	
<b>A2</b>	<b>Auteur(s)</b> Nom, Prénom, Affiliation	
<b>A3</b>	<b>Discipline/Champ d'étude</b>	
<b>A4</b>	<b>Dates de l'enquête</b> du projet de recherche à la publication majeure	JJ.MM.AAAA
<b>A5</b>	<b>Source(s) de financement</b>	
<b>A6</b>	<b>Résumé</b> précisant les buts, finalités, objets, problématiques, etc. de l'enquête <i>Vous êtes invité à copier un extrait de votre projet</i>	

<https://docs.google.com/document/d/1PI4O0kE3DVMfAHfGluhA4PHvFvS8Pm50/edit>

---

# Panorama de la collection “Enquêtes électorales”

**French Election  
Study**

**Données de  
recherche**

PEOPLE  
PELMEL

**Données de recherche du  
CEVIPOF**








Panel électoral français  
Enquête post-électorale  
française  
Baromètre Politique Français  
2006-2007

**Enquête Électorale française 2017**

Base de données comparatives des enquêtes  
électorales françaises

---

# Focus ENEF 2017 - Documentation

	<p><b>Documentation_ENEF2017_V1.pdf</b> Adobe PDF - 1.0 Mo Publié 30 oct. 2024 38 téléchargements MD5: cbf...c82 </p> <p><b>Documentation</b></p>
	<p><b>ENEF2017_V1aV18_Dictionnaire.zip</b> Archive ZIP - 123.2 Ko Publié 6 nov. 2024 22 téléchargements MD5: 130...935 </p> <p>Dictionnaire variables</p> <p><b>Documentation</b></p>
	<p><b>fr_cdsp_ddi_enef2017_v1.xml</b> XML - 1.0 Mo Publié 30 oct. 2024 17 téléchargements MD5: a67...63c </p> <p><b>Documentation</b> <b>XML</b></p>
	<p><b>Questionnaire_ENEF2017_V1.docx</b> MS Word - 197.8 Ko Publié 30 oct. 2024 25 téléchargements MD5: bf5...899 </p> <p><b>Documentation</b> <b>Questionnaire</b></p>
	<p><b>RapportTechnique_ENEF2017_V1.pdf</b> Adobe PDF - 746.9 Ko Publié 23 oct. 2024 25 téléchargements MD5: 9a5...0c0 </p> <p><b>Documentation</b></p>

Codebook

Liste des  
variables

DDI XML

Questionnaire

Rapport  
pondération

# Focus ENEF 2017 - Données



fr\_cdsp\_ddi\_enef2017\_v1-1.tab

Données tabulaires - 36.9 Mo

Publié 30 oct. 2024

7 téléchargements

294 Variables, 24369 Observations UNF:6:H+Me...qWg==

Données STATA



fr\_cdsp\_ddi\_enef2017\_v1.tab

Données tabulaires - 23.7 Mo

Publié 30 oct. 2024

6 téléchargements

294 Variables, 24369 Observations UNF:6:H+Me...qWg==

Données SPSS



fr\_cdsp\_ddi\_enef2017\_v1\_csv

Archive ZIP - 4.3 Mo

Publié 30 oct. 2024

8 téléchargements

MD5: fe9...360

CSV Données



fr\_cdsp\_ddi\_enef2017\_v1\_sas

Archive ZIP - 5.6 Mo

Publié 30 oct. 2024

5 téléchargements

MD5: fb1...945

Données SAS



fr\_cdsp\_ddi\_enef2017\_v1\_verbatims.zip

Archive ZIP - 275.2 Ko

Publié 30 oct. 2024

5 téléchargements

MD5: 517...7e9

CSV Données

**TO ALL**

**ASK ENJ1 AND ENJ2 ON THE SAME SCREEN**

ENJ1

**A votre avis, quel est le problème LE PLUS IMPORTANT pour la France aujourd'hui ?**

*Votre réponse doit être courte et une seule idée doit être exprimée*

**OPEN ENDED QUESTION**

*Notez en clair*

**TO ALL**

ENJ2

**En second ? A votre avis, quel est LE SECOND problème le plus important pour la France aujourd'hui ?**

**OPEN ENDED QUESTION**

*Notez en clair*

**ASK ENJ3 IF AT LEAST ONE ANSWER AT ENJ1 OR ENJ2**

ENJ3A / ENJ3B

**Selon vous quels partis politiques sont les plus aptes à s'occuper du/des problème(s) que vous venez de noter précédemment ?**

*Une seule réponse possible par colonne*

**DISPLAY COLUMN ENJ3A IF AN ANSWER AT ENJ1**

**DISPLAY COLUMN ENJ3B IF AN ANSWER AT ENJ2**

**SINGLE ANSWER per column**

	column ENJ3A	column ENJ3B
	Le problème le plus important	Le second problème :
	DISPLAY HERE ANSWER AT ENJ1	DISPLAY HERE ANSWER AT ENJ2
Lutte ouvrière ou NPA (Nouveau parti anticapitaliste)	1	1
Le Parti communiste	2	2
Le Parti de Gauche	3	3
Nouvelle Donne	4	4
Le Parti Socialiste	6	6
Europe Ecologie – Les Verts	5	5
Le MoDem	7	7
L'Union des Démocrates et Indépendants (UDI)	8	8
Les Républicains (ex-UMP)	9	9
Le Front National	10	10
Un autre parti	11	11
Aucun parti n'est capable de s'en occuper	97	97
Je ne sais pas	99	99

Réponse ouverte

# Curation et enrichissement

## Pseudonymisation

- PCS, Région de résidence ou inscription...
- Données moins détaillées disponible sur demande

## Documentation et enrichissement

- Données plus lisible
- Différenciation des NA

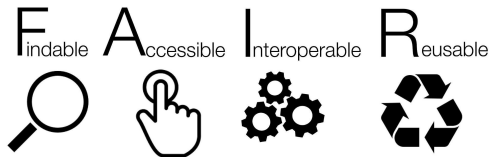
# AQ10BIS: Proximité avec le parti politique concerné			
Information	[Type= discrete] [Format=numeric] [Range= 1-9999] [Missing=*]		
Statistics [NW/ W]	[Valid=18677 /-] [Invalid=5692 /-]		
Universe	Si l'interviewé a répondu qu'il y a un parti politique duquel il se sent proche, ou moins éloigné que d'autres (c'est-à-dire si AQ8 = 1 ou AQ9 = 1).		
Literal question	Diriez-vous que vous êtes très proche, assez proche, pas très proche ou pas proche du tout de ce parti ?		
Value	Label	Cases	Percentage
1	Très proche	1983	10.6%
2	Assez proche	11659	62.4%
3	Pas très proche	4206	22.5%
4	Pas proche du tout	829	4.4%
6666	Non concerné(e)	5689	
9999	N'a pas répondu	3	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*

---

# Publications à venir

- Enquête électorale française 2019 : 5 vagues  
(<https://www.sciencespo.fr/cevipof/fr/etudes-enquetes/enquete-electorale-francaise-2019/>)
- Enquête électorale française 2022 : 12 vagues  
(<https://www.sciencespo.fr/cevipof/fr/etudes-enquetes/enquete-electorale-francaise-2022-electio ns-legislatives-2022/>)
- PEF 2012 Présidoscopie : 12 vagues



with



## Implementing the FAIR principles with the metadata standard Data Documentation Initiative (DDI)

Funded by the French National Research Agency

Coordination: Alina Danciu

Work Packages:

- WP1: The R of FAIR (Evaluation criterias)
- WP2: Relying on the Data Documentation Initiative (DDI) for more FAIR data (training materials)
- **WP3: Upscaling metadata using AI (harmonisation)**





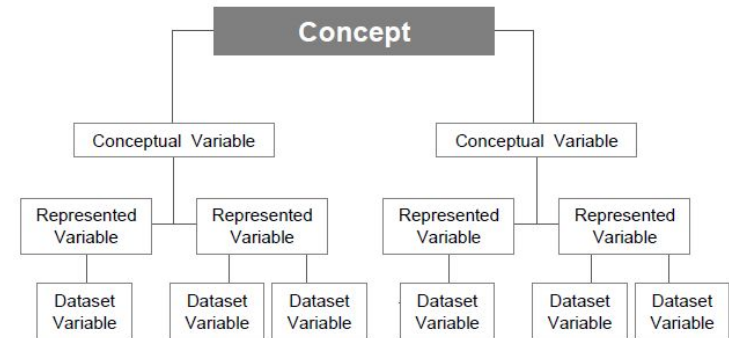
---

# WP3: Upscaling metadata using AI

## Improving data comparability and reusability across heterogeneous datasets

Establishing a machine-actionable **harmonisation protocol** based on the **DDI-Lifecycle “variable cascade” architecture**

- Automating variable linkages to ELSST concepts using Machine Learning models
- Testing the model with external metadata (FSD, Constances)



<https://doi.org/10.5281/zenodo.6646078>

Sharing the harmonised metadata in a **question bank prototype** built by the CDSP developers.

## Résultats de la recherche

politique

Séries ▾

Études ▾

Questions ▾

Rechercher

Affichage de l'élément 1 à 10 sur 92 éléments (filtré à partir de 2,993 éléments au total)

Est-ce que vous vous intéressez à la **politique** ?

Enquête : Enquête annuelle - vague 7 (2019)

Nom de la variable : ea19\_11

Label interne : Intérêt pour la politique

Modalités

Vous participez aux activités de cette association... : -Parti **politique**

Enquête : Enquête annuelle - vague 8 (2021)

Nom de la variable : ea21\_f7bis\_12

Label interne : Participation aux activités d'une association : parti politique

Modalités

Est-ce que vous vous intéressez à la **politique** ?

Enquête : Enquête annuelle - vague 8 (2021)

Nom de la variable : ea21\_11

Label interne : Intérêt pour la politique

Modalités

Est-ce que vous vous intéressez à la **politique** ?

Enquête : Enquête annuelle - vague 10 (2023)

Nom de la variable : ea23\_11

Label interne : Intérêt pour la politique

Modalités

Vous participez aux activités de cette association... : -Parti **politique**

Enquête : Enquête annuelle - vague 9 (2022)

Nom de la variable : ea22\_f7bis\_12

Label interne : Participation aux activités d'une association : parti politique

Modalités

# Question bank preview

- Import: DDI-C xml and csv
- Harmonisation: automatic clustering **identical and equivalent questions** (represented variable & conceptual variable items generation)
- Search: Open Source Technology

## Import de fichier XML

[Créer une nouvelle série](#)

### Balises attendues pour le XML

- var
  - catValu
  - labl
  - catgr
  - qstDLE
  - universe
- notes
- IDNo
- tit

Assurez-vous que ces balises sont présentes dans votre fichier XML pour garantir une structure valide.

### Uploader un fichier XML

Sélectionnez une série: ▾

Sélectionnez un fichier XML:

[Choisir un fichier](#) | Aucun fichier choisi

[Envoyer le fichier XML](#)



elasticsearch



# ML experiment

## Creating concepts linkage to questions

- Time-consuming manual tagging
- Subjective human gaps
- Difficulty in identifying relevant concepts from over 3,000 options

### Objective

To automate the attribution of **European Language Social Science Thesaurus** concepts to the *Étude Longitudinale par Internet Pour les Sciences Sociales* questions



+100 surveys  
+30k variables



# ML experiment

Question: Quels sont les avantages et les inconvénients du télétravail selon vous ?

Top 1 Predicted Concept: NUMÉRISATION with confidence 0.00503162806853652  
Top 2 Predicted Concept: BUREAUCRATIE with confidence 0.004637261852622032  
Top 3 Predicted Concept: CONTRATS DE TRAVAIL with confidence 0.004574696999043226

Question: Pensez-vous que les impôts vont augmenter dans les prochaines années ?

Top 1 Predicted Concept: IMPÔT SUR LE REVENU with confidence 0.006328627932816744  
Top 2 Predicted Concept: DIVORCE with confidence 0.005007258616387844  
Top 3 Predicted Concept: ALLOCATIONS DE CHÔMAGE with confidence 0.004978996701538563

Question: Quelle est votre opinion sur le système éducatif actuel ?

Top 1 Predicted Concept: ACCÈS À L'ÉDUCATION with confidence 0.00794867891818285  
Top 2 Predicted Concept: ÉDUCATION with confidence 0.006567481905221939  
Top 3 Predicted Concept: ENSEIGNEMENT with confidence 0.006182760465890169

Question: Comment évaluez-vous les efforts du gouvernement pour lutter contre le changement climatique ?

Top 1 Predicted Concept: CHANGEMENTS CLIMATIQUES with confidence 0.007203781511634588  
Top 2 Predicted Concept: RÉCHAUFFEMENT PLANÉTAIRE with confidence 0.006725126411765814  
Top 3 Predicted Concept: PRIX DE L'ÉNERGIE with confidence 0.004714252892881632

Question: Quelle est votre opinion sur l'utilisation des énergies renouvelables ?

Top 1 Predicted Concept: ÉNERGIE with confidence 0.00679916562512517  
Top 2 Predicted Concept: TECHNOLOGIE ET INNOVATION with confidence 0.005647766403853893  
Top 3 Predicted Concept: UTILISATION D'INTERNET with confidence 0.005034273490309715

Question: Pensez-vous que l'intelligence artificielle va transformer le marché du travail ?

Top 1 Predicted Concept: INTELLIGENCE ARTIFICIELLE with confidence 0.008562188595533371  
Top 2 Predicted Concept: NUMÉRISATION with confidence 0.0041154227219522  
Top 3 Predicted Concept: ACCÈS À LA TECHNOLOGIE DE L'INFORMATION ET DES COMMUNICATIONS with confidence 0.0041154227219522

## Method

Generation of additional training data using ChatGPT to filter concepts and generate questions related to the concepts.

## Last version of the model (97%)

Suitable for predicting multiple concepts, open to improvement



# BANQUE DE DONNÉES DU CDSP

**Merci pour votre attention**

[info.cdsp@sciencespo.fr](mailto:info.cdsp@sciencespo.fr)

<https://www.sciencespo.fr/cdsp/fr/>

<https://data.sciencespo.fr/dataverse/cdsp>

